14th International Conference on Information Fusion
Chicago, Illinois, USA, July 5-8, 2011

Multimodal Sensor Fusion for Personnel Detection

Xin Jin Shalabh Gupta Asok Ray
Department of Mechanical Engineering
The Pennsylvania State University
University Park, PA 16802, USA
Email: xuj103@psu.edu, szg107@psu.edu, axr2@psu.edu

Thyagaraju Damarla
Networked Sensing and Fusion Branch
U.S. Army Research Laboratory
Adelphi, MD 20783, USA
Email: thyagaraju.damarla@us.army.mil

***Abstract*—Unattended ground sensors (UGS) are widely used to monitor human activities, such as pedestrian motion and detection of intruders in a secure region. Efficacy of UGS systems is often limited by high false alarm rates, possibly due to inadequacies of the underlying algorithms and limitations of onboard computation. This paper presents a symbolic method of feature extraction and sensor fusion, which is built upon the principles of wavelet transform and probabilistic finite state automata (PFSA). The relational dependencies among heterogeneous sensors are modeled by cross-PFSA, from which low-dimensional feature vectors are generated for pattern classification in real time. The proposed method has been validated on data sets of seismic and passive infrared (PIR) sensors for target detection and classification. The proposed method has the advantages of fast execution time and low memory requirements and is potentially well-suited for real-time implementation with onboard UGS systems.**

***Index Terms*—Personnel detection, multimodal sensor fusion, feature extraction, seismic sensor, PIR sensor**

I. Introduction

Unattended ground sensors (UGS) are widely used in industrial monitoring and military operations. Such UGS are usually lightweight devices that automatically monitor the local activities in-situ, and transfer target detection and classification reports to some higher level processing center. Commercially available UGS systems make use of multiple sensing modalities (e.g., acoustic, seismic, passive infrared, magnetic, electrostatic, and video). Efficacy of UGS systems is often limited by high false alarm rates because the onboard data processing algorithms may not be able to correctly discriminate different types of targets (e.g., humans from animals) [1]. Power consumption is a critical consideration in UGS systems. Therefore, power-efficient sensing modalities, low-power signal processing algorithms, and efficient methods for exchanging information between the UGS nodes are needed [2].

In a personnel detection problem, the targets usually include human, vehicles, and animals. Discriminating human footstep signals from other targets and noise sources is a challenging problem, because the signal to noise ratio (SNR) of footsteps decreases rapidly with the distance between the sensor and the pedestrian. Furthermore, the footstep signals may vary significantly for different persons and environments.

Seismic sensors are widely used for personnel detection, because they are relatively less sensitive to Doppler effects environment variations as compared to acoustic sensors [3]. Current personnel detection methods using seismic signals can be classified into three categories, namely, time domain methods [4], frequency domain methods [5], and time-frequency domain methods [3], [6]. Recent research has relied on time-frequency domain methods, such as wavelet transform-based methods. Passive Infrared (PIR) sensors are widely used for motion detection, and are well-suited for UGS systems due to low power consumption. PIR sensors have been reported for moving targets detection and localization [7]; however, similar effort for target classification has not been reported in open literature, although PIR sensor signals also contain discriminative information in the time-frequency domain.

Collaborative target detection and classification using multimodal sensor fusion would increase the overall performance because the heterogeneous sensors can complement each other. Sensor fusion can be implemented at different levels: data-level fusion, feature-level fusion, and decision-level fusion. Kalman filter is widely used for data-level fusion; Dempster-Shafer evidence theory and Bayesian network are widely used for decision-level fusion [8] [9]. Data-level fusion has the least information loss, but it may be computationally expensive and vulnerable to sensor degradation. Some of these concerns can be alleviated by decision-level fusion, in which detection/classification is performed at the data-level and then the decisions are combined from individual sensors. In principle, decision-level fusion is suboptimal since if a target is not detected by all sensors, it will not experience the full benefits of fusion [10].

This paper introduces a feature-level fusion method to address these issues. Symbolic Dynamic Filtering (SDF) is a data-driven feature extraction tool built upon the concepts of Symbolic Dynamics and Probabilistic Finite State Automata (PFSA) [11] [12]. In SDF, the sensor data are first partitioned into symbol sequences, and then PFSA are constructed as the representation of the underlying dynamics in the data. A feature-level fusion approach built under the framework of SDF has been proposed in [13] for fault diagnosis in aircraft engine. The time series data from different sensors are partitioned into symbol sequences from which the cross-PFSA, called cross D-Markov machine [13] (denoted as 'xD-Markov machine' in the sequel), is constructed. However, the performance of this method may degrade significantly if the SNR decreases. For analysis of noisy sensor data, this paper extends the concept of xD-Markov machines by introducing

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JUL 2011		**00-00-2011 to 00-00-2011**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Multimodal Sensor Fusion for Personnel Detection	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **U.S. Army Research Laboratory,Networked Sensing and Fusion Branch,Adelphi,MD,20783-1197**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Presented at the 14th International Conference on Information Fusion held in Chicago, IL on 5-8 July 2011. Sponsored in part by Office of Naval Research and U.S. Army Research Laboratory.

14. ABSTRACT
Unattended ground sensors (UGS) are widely used to monitor human activities, such as pedestrian motion and detection of intruders in a secure region. Efficacy of UGS systems is often limited by high false alarm rates, possibly due to inadequacies of the underlying algorithms and limitations of onboard computation. This paper presents a symbolic method of feature extraction and sensor fusion, which is built upon the principles of wavelet transform and probabilistic finite state automata (PFSA). The relational dependencies among heterogeneous sensors are modeled by cross-PFSA, from which low-dimensional feature vectors are generated for pattern classification in real time. The proposed method has been validated on data sets of seismic and passive infrared (PIR) sensors for target detection and classification. The proposed method has the advantages of fast execution time and low memory requirements and is potentially well-suited for real-time implementation with onboard UGS systems.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**8**	

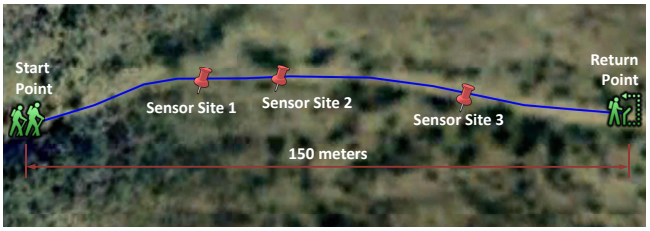


Figure 1. An illustration of the test scenario with three sensor sites

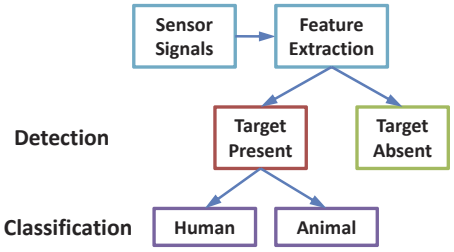


Figure 2. Tree structure formulation of the detection & classification problem

wavelet surface partitioning [14] as an alternative to time series partitioning in the original concept in [13]. In the proposed method, images of wavelet-transformed time series are partitioned for conversion into symbol sequences. Subsequently, xD-Markov machines are constructed from symbol sequences of heterogeneous sensors to compress the pertinent information into low-dimensional statistical patterns. The proposed feature extraction algorithm mitigates the detrimental effects of spurious noise by using wavelet analysis, captures the essential signatures from the time-frequency domain of the signals, and generates low-dimensional feature vectors for pattern classification.

The proposed method is validated on the data collected from seismic sensors and PIR sensors for the purpose of personnel detection in border area. Performance of information fusion from seismic and PIR sensors is compared with the results obtained from single-modal sensors.

II. Problem Description and Formulation

The problem at hand is to detect and classify different targets (e.g., humans and animals), where seismic and PIR sensors are used to capture the respective characteristic signatures. For example, in the movement of a human or an animal across the ground, oscillatory motions of the body appendages provide the respective characteristic signatures.

The seismic and PIR sensor data, used in this analysis, were collected on multiple days from test fields on a wash (i.e., the dry bed of an intermittent creek) and at a choke point (i.e., a place where the targets are forced to go due to terrain difficulties). During multiple field tests, sensor data were collected for several scenarios that consisted of targets walking along an approximately 150 meters long trail, and returning along the same trail to the starting point. Figure 1 illustrates a typical data collection scenario.

The targets consisted of people (e.g., male and female humans) and animals (e.g., donkeys, mules, and horses). The humans walked alone and in groups with and without backpacks; the animals were led by their human handlers and they made runs with and without payloads. There were three sensor sites, each equipped with acoustic and seismic sensors. The seismic sensors were buried approximately 15 cm deep underneath the soil surface, and the PIR sensors were collocated with the respective seismic sensors. All targets passed by the sensor sites at a distance of approximately 5 m. Signals from both sensors were acquired at a sampling frequency of 10 kHz.

The tree structure in Fig. 2 shows how the detection and classification problem is formulated. In the detection stage, the pattern classifier detects the presence of a moving target against the null hypothesis of no target present; in the classification stage, the pattern classifier discriminates among different targets. While the detection system should be robust to satisfy the specifications of false alarm rates, the classification system must be sufficiently sensitive to discriminate between different classes of targets with high fidelity. In this context, feature extraction plays an important role in target detection and classification, because the performance of the classifier largely depends on the quality of the extracted features.

III. Semantic Framework of Sensor Fusion

A (three-layered) hierarchical semantic framework is presented in this paper for the purpose of multi-sensor data interpretation and fusion. In this framework, patterns discovered from individual sensors are called atomic patterns (AP), while patterns discovered from the relational dependency between two sensors are called relational patterns (RP) [13].

Let $\mathbb{L} = \{\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_N\}$ be the universal set of *atomic* patterns. The *atomic* pattern library $\mathbb{L}$ is set of modal footprints identified from individual sensing modalities for targets of different classes. Given the atomic pattern library, a popular framework for addressing information fusion is what is called the *set-theoretic* approach. In this framework, higher level patterns or contexts are modeled as subsets of $\mathbb{L}$. Thus a composite pattern, resulting from fusion of atomic patterns, is a collection of atomic patterns from $\mathbb{L}$ and the resulting library of composite patterns is a subset of the power set of the atomic pattern library, i.e., $\mathbb{L}^* \subseteq 2^{\mathbb{L}}$. However, a disadvantage of this approach is that it considers only modal footprints for constructing composite patterns as a *bag of atomic patterns*; relational dependencies between patterns are disregarded.

Since the relational dependencies cannot be ignored in many practical problems, a hierarchical semantic framework for multi-sensor data interpretation and fusion is proposed in this paper, which involves a common approach to information fusion at different layers of the hierarchy and to include relational dependencies for composite pattern representation. Thus, the middle layer deals with the relational dependencies among atomic patterns, where relationships are modeled as the cross-dependencies among sensor data streams from different sensors. These cross-dependencies are discovered via relational PFSA that essentially capture the dynamics of state

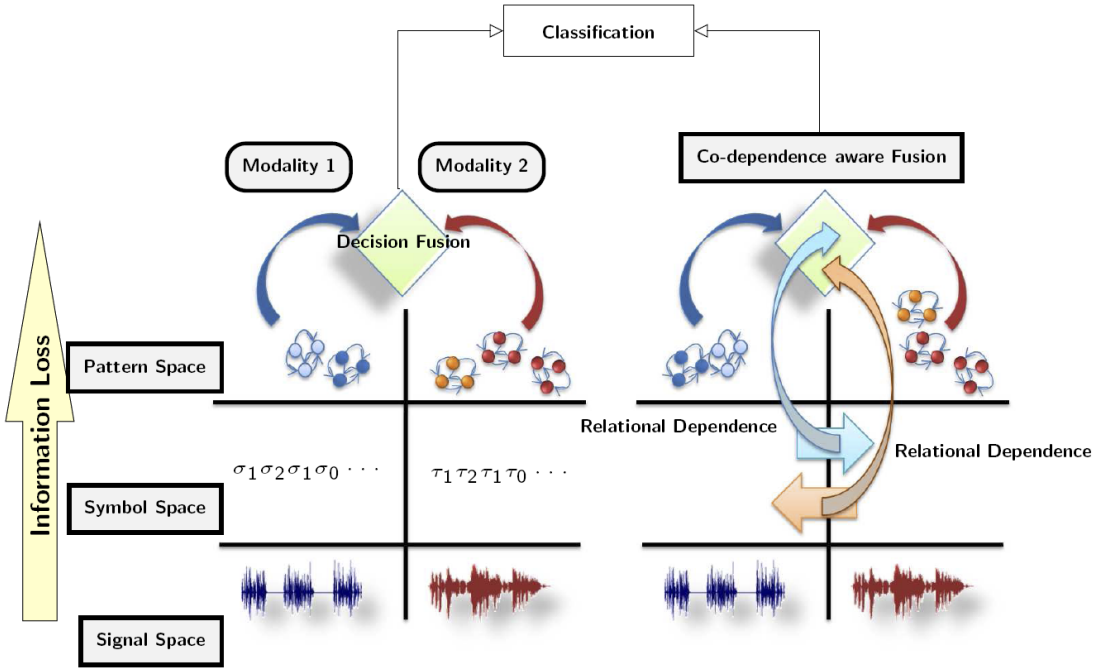


Figure 3. Comparison of the set-theoretic fusion (left) and the proposed semantic fusion of multimodal sensors (right)

transition in one symbol sequence (e.g., obtained from one sensor) corresponding to a symbol appearance in the second symbol sequence (e.g., obtained from another sensor). Loose time-synchronization between sensor observations should be adequate for this purpose. Symbol-level cross-dependencies among modalities are exploited to mitigate information loss.

Finally, the top layer consists of higher level composite patterns that is represented as digraphs where the atomic patterns (AP) are modeled as nodes and dependencies between nodes are modeled as relational patterns (RP). An illustrative example in Fig. 3 compares the set-theoretic fusion (left) with the proposed co-dependence aware fusion (right). The definition of the composite pattern is as follows.

Definition 3.1 (Composite pattern representation): Let $\mathbb{L} = \{\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_N\}$ be the atomic pattern library and let $\mathbb{L}^* \subseteq 2^{\mathbb{L}}$ be the set of allowable primitives for a class. Then, a composite pattern library $\mathbb{G}^r = \{\mathcal{G}_1^r, \mathcal{G}_2^r, \ldots, \mathcal{G}_M^r\}$ where the i^{th} composite pattern $\mathcal{G}_i^r$ is a digraph $\mathcal{G}_i^r = (\mathcal{L}_{V_i}, \mathcal{E}_{V_i})$; $\mathcal{L}_{V_i} \subseteq \mathbb{L}$ with the index set $V_i \subseteq \{1, 2, \ldots, N\}$ and $\mathcal{E}_i = \{\mathcal{R}_{jk}, ...\}$ is a set of *relational* PFSAs.

The relational PFSAs are discovered by cross D-Markov machine [13] construction to determine the respective cross-dependence; the algorithm is described in Section III-C. Absence of a directed edge in composite pattern digraph would be represented by a single state machine for relational PFSA, which implies the lack of prediction capability of a target state by the parent state.

A. Sensor Signal Conditioning and Transformation

This section presents the procedure for generation of wavelet coefficient, i.e., an image in the scale-shift domain, denoted as 'wavelet image' in the sequel, from observed sensor time series for construction of symbolic representations of the underlying dynamics. In this SDF-based procedure, a crucial step is partitioning of the phase space for symbol sequence generation. Various partitioning techniques have been reported in literature, and a brief review is given in [14].

In wavelet-based partitioning, time series are first transformed to wavelet domain, where wavelet coefficients are generated at different time shifts and scales. The choice of the wavelet basis function and wavelet scales depends on the time-frequency characteristics of individual signals.

For every wavelet, there exists a certain frequency called the center frequency F_c that has the maximum modulus in the Fourier transform of the wavelet. The pseudo-frequency f_p of the wavelet at a particular scale α is given by the following formula:

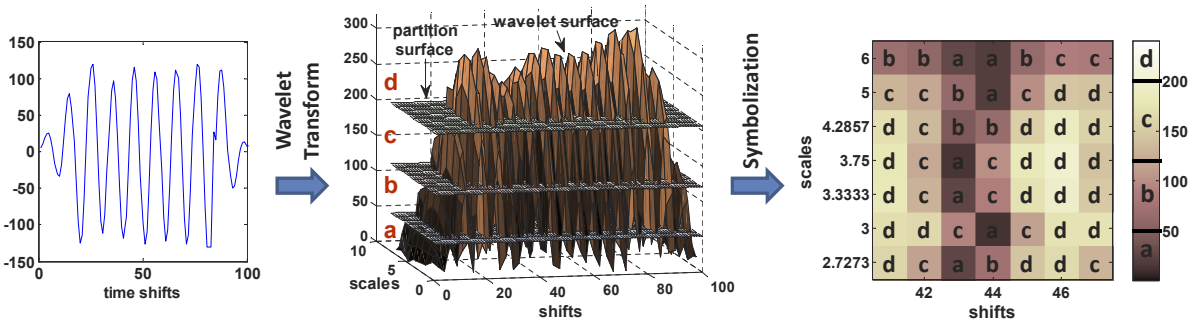
$$f_p = \frac{F_c}{\alpha\ \Delta t}, \tag{1}$$

where Δt is the sampling interval. Then the scales can be calculated as follows:

$$\alpha^i = \frac{F_c}{f_p^i\ \Delta t} \tag{2}$$

where $i = 1, 2, ...,$ and f_p^i are the frequencies that can be obtained by choosing the locally dominant frequencies in the Fourier transform.

Figure 4 shows an illustrative example of transformation of the time series in Fig. 4(a) to the two-dimensional wavelet image in Fig. 4(b); the amplitudes of the wavelet coefficients over the scale-shift domain are plotted as a surface. Subsequently, symbolization of this wavelet surface leads to the formation of a symbolic image as shown in Fig. 4(c).



(a) Sensor time series data (b) Partition of the wavelet coefficients (c) Symbolized wavelet image (a section)

Figure 4. Symbol image generation via wavelet transform of the sensor time series data and partition of the wavelet surface in ordinate direction

B. Symbolization of Wavelet Surface Profiles

This section presents partitioning of the wavelet surface profile, as shown in Fig. 4(b), which is generated by the coefficients over the two-dimensional scale-shift domain, for construction of the symbolic image in Fig. 4(c). The $x-y$ coordinates of the wavelet surface profiles denote the shifts and the scales respectively, and the z-coordinate denotes the pixel values of wavelet coefficients (i.e., the surface height).

Definition 3.2: (*Wavelet Surface Profile*) Let $\mathcal{H} \triangleq \{(i,j) : i,j \in \mathbb{N},\ 1 \leq i \leq m,\ 1 \leq j \leq n\}$ be the set of coordinates consisting of $(m \times n)$ pixels denoting the scale-shift data points. Let $\mathcal{R}$ denote the interval that spans the range of wavelet coefficient amplitudes. Then, a wavelet surface profile is defined as

$$\mathcal{S} : \mathcal{H} \rightarrow \mathcal{R} \tag{3}$$

Definition 3.3: (*Symbolization*) Given the symbol alphabet Σ, let the partitioning of the interval $\mathcal{R}$ be defined by a map $P : \mathcal{R} \rightarrow \Sigma$. Then, the symbolization of a wavelet surface profile is defined by a map $\mathcal{S}_\Sigma \equiv P \circ \mathcal{S}$ such that

$$\mathcal{S}_\Sigma : \mathcal{H} \rightarrow \Sigma \tag{4}$$

that labels each pixel of the image to a symbol in Σ.

The wavelet surface profiles are partitioned such that the ordinates between the maximum and minimum of the coefficients along the z-axis are divided into regions by different planes parallel to the $x-y$ plane. For example, if the alphabet is chosen as $\Sigma = \{a,b,c,d\}$, i.e., $|\Sigma| = 4$, then three partitioning planes divide the ordinate (i.e., z-axis) of the surface profile into four mutually exclusive and exhaustive regions, as shown in Figure 4 (b). These disjoint regions form a partition, where each region is labeled with one symbol from the alphabet Σ. If the intensity of a pixel is located in a particular region, then it is coded with the symbol associated with that region. As such, a symbol from the alphabet Σ is assigned to each pixel corresponding to the region where its intensity falls. Thus, the two-dimensional array of symbols, called *symbol image*, is generated from the wavelet surface profile, as shown in Figure 4 (c).

The surface profiles are partitioned by using either the maximum entropy partitioning (MEP) or the uniform partitioning (UP) methods [14]. If the partitioning planes are separated by equal-sized intervals, then the partition is called the *uniform partitioning* (UP). Intuitively, it is more reasonable if the information-rich regions of a data set are partitioned finer and those with sparse information are partitioned coarser. To achieve this objective, the MEP method has been adopted such that the entropy of the generated symbols is maximized. In general, the choice of alphabet size depends on specific data set. The partitioning of wavelet surface profiles to generate symbolic representations enables robust feature extraction, and symbolization also significantly reduces the memory requirements. For the purpose of pattern classification, the reference data set is partitioned with alphabet size $|\Sigma|$ and is subsequently kept constant. In other words, the structure of the partition is fixed at the reference condition and this partition serves as the reference frame for subsequent data analysis [11].

C. Construction of PFSA for Feature Extraction

This section presents construction of a probabilistic finite state automaton (PFSA) for feature extraction based on the symbol image generated from a wavelet surface profile.

For analysis of (one-dimensional) time series, a PFSA is constructed such that its states represent different combinations of blocks of symbols on the symbol sequence. The edges connecting these states represent the transition probabilities between these blocks [11]. Therefore, for analysis of (one dimensional) time series, the 'states' denote all possible symbol blocks (i.e., words) within a window of certain length. Let us now extend the notion of 'states' on a two-dimensional domain for analysis of wavelet surface profiles.

The concept of D-Markov machine has been introduced by the authors in their previous publications under the framework of Symbol Dynamic Filtering (SDF) [11], [14] to extract information from symbol sequences/images which are generated from single sensors. In this paper, a generalization of the D-Markov machine is proposed, called xD-Markov machine, which captures the symbol level cross-dependence. The D-Markov machine is a special case of the xD-Markov machine in the following sense: when both symbol sequences are the same, the relational patterns are essentially the atomic patterns corresponding to the symbol sequence; i.e., xD-Markov machine reduces to a simple D-Markov machine. The feature vectors extracted from xD-Markov machine with two

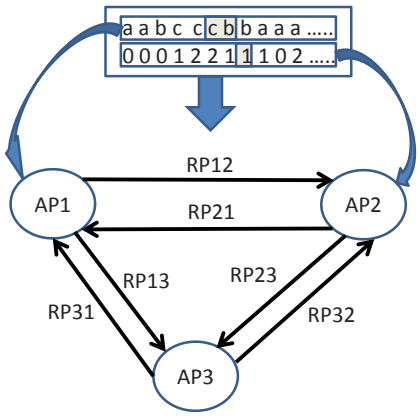


Figure 5. Composite Pattern Digraph

same symbol sequences (i.e., from D-Markov machine) are called Atomic Pattern (AP), and those extracted from xD-Markov machine with two different symbol sequences are called Relational Pattern (RP). The digraph representation of AP and RP is illustrated in Fig. 5.

The xD-Markov machines are constructed based on two symbol sequences $\{\mathbf{s}_1\}$ and $\{\mathbf{s}_2\}$ obtained from two different sensors (possibly heterogeneous) to capture the symbol level cross-dependence. Conversion of the two-dimensional symbol image to one-dimensional symbol sequence while retaining the pertinent information is a difficult problem. One simple way for the conversion is to stack each row in the symbol image one after another. However, this method only works if the two symbol sequences $\{\mathbf{s}_1\}$ and $\{\mathbf{s}_2\}$ have the same number of scales in wavelet transform; otherwise their length of the generated symbol sequences will not be equal. More importantly, this method may suffer information loss in the frequency domain by only looking for relational dependency at the similar frequency bands in $\{\mathbf{s}_1\}$ and $\{\mathbf{s}_2\}$. It is highly possible that the low frequency component in $\mathcal{H}_1$ has stronger correlation with the high frequency component in $\mathcal{H}_2$.

To avoid these issues, a new method for converting symbol images from two sensors to symbol sequences for discovery of relational dependency is proposed. The key idea is to exhaustively find all possible combination of rows between the two symbol images. A formal definition is as follows:

Definition 3.4 (Conversion): Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be the wavelet coefficients (images) of sensor 1 and sensor 2, consisting of $m_1 \times n$ and $m_2 \times n$ pixels, respectively. Let $\mathcal{W}_i^j \subseteq \mathcal{H}_i$ be the window that covers the jth scale in $\mathcal{H}_i$. Then the two symbol sequences $\{\mathbf{s}_1\}$ and $\{\mathbf{s}_2\}$ are defined as

$$\{\mathbf{s}_1\} = [\underbrace{\mathcal{S}_\Sigma(\mathcal{W}_1^1)\ldots\mathcal{S}_\Sigma(\mathcal{W}_1^1)}_{m_2}, \cdots, \underbrace{\mathcal{S}_\Sigma(\mathcal{W}_1^{m_1})\ldots\mathcal{S}_\Sigma(\mathcal{W}_1^{m_1})}_{m_2}]$$

$$\{\mathbf{s}_2\} = [\underbrace{\mathcal{S}_\Sigma(\mathcal{W}_2^1)\ldots\mathcal{S}_\Sigma(\mathcal{W}_2^{m_2}), \cdots, \mathcal{S}_\Sigma(\mathcal{W}_2^1)\ldots\mathcal{S}_\Sigma(\mathcal{W}_2^{m_2})}_{m_1}]$$

By implementing the conversion procedure as defined above, the wavelet images $\mathcal{H}_1$ of $(m_1 \times n)$ pixels and $\mathcal{H}_2$ of $(m_2 \times n)$ are converted to one-dimensional symbol sequences

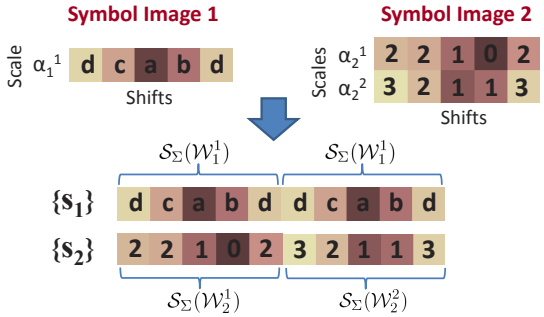


Figure 6. An illustration of converting symbol images to symbol sequences

of the same length $(m_1 \times m_2 \times n)$. An illustration is given in Fig. 6, where $m_1 = 1$, $m_2 = 2$, $n = 5$ and $|\Sigma_1| = |\Sigma_2| = 4$. A formal definition of the xD-Markov machine is as follows:

Definition 3.5 (xD-Markov): Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be the PFSAs corresponding to symbol sequences $\{\mathbf{s}_1\}$ and $\{\mathbf{s}_2\}$ respectively. Then a xD-Markov machine is defined as a 4-tuple $\mathcal{M}_{1\to2} \triangleq (\mathcal{Q}_1, \Sigma_2, \delta_{12}, \tilde{\Pi}_{12})$ such that:

- $\Sigma_2 = \{\sigma_0, ..., \sigma_{|\Sigma_2|-1}\}$ is the alphabet set of symbol sequence $\{\mathbf{s}_2\}$
- $\mathcal{Q}_1 = \{q_1, q_2, \ldots, q_{|\Sigma|_1^{D_1}}\}$ is the state set corresponding to symbol sequence $\{\mathbf{s}_1\}$, where D_1 is the depth for $\{\mathbf{s}_1\}$
- $\delta_{12} : \mathcal{Q}_1 \times \Sigma_2 \to \mathcal{Q}_1$ is the state transition mapping that maps the transition in symbol sequence $\{\mathbf{s}_1\}$ from one state to another upon arrival of a symbol in $\{\mathbf{s}_2\}$
- $\tilde{\Pi}_{12}$ is the symbol generation matrix of size $|\mathcal{Q}_1| \times |\Sigma_2|$; the ij element of $\tilde{\Pi}_{12}$ denotes the probability of finding j^{th} symbol in $\{\mathbf{s}_2\}$ while making a transition from i^{th} state in the symbol sequence $\{\mathbf{s}_1\}$

In practice, $\tilde{\Pi}_{12}$ is reshaped into a vector $\mathbf{p}_{12}$ of length $|\mathcal{Q}_1| \times |\Sigma_2|$ and is treated as the extracted feature vector that is a representation of the relational dependence between $\{\mathbf{s}_1\}$ and $\{\mathbf{s}_2\}$. This feature vector is called a Relational Pattern (RP). The xD-Markov machine $\mathcal{M}_{2\to1}$ and the corresponding feature vector $\mathbf{p}_{21}$ are defined similarly. Fig. 5 schematically describes the basic concept of the xD-Markov machine. Note, a RP between two symbol sequences is not symmetric; therefore, RPs need to be identified for both directions. If $\{\mathbf{s}_1\}$ and $\{\mathbf{s}_2\}$ are the same, then the xD-Markov machines $\mathcal{M}_{1\to1}$ and $\mathcal{M}_{2\to2}$ reduce to the simple D-Markov machine, and the feature vector obtained from $\tilde{\Pi}_{11}$ or $\tilde{\Pi}_{22}$ is called an Atomic Pattern (AP).

The set-theoretic approach falls at one end of the spectrum of information fusion; here all relationships are excluded and any fusion is solely done in the decision-theoretic sense where the presence (or absence) of one or more footprints can be used to estimate the probability of the fault class under consideration. The other end of the spectrum is to fuse data at the lowest level and construct machines (PFSAs) working in the product space of all sensors. This approach would be able extract modal dependencies before they are lost when constructing separate machines for individual sensor or modalities. But working in the product space has the danger of state space explosion especially when the sensors and sensing

modalities can be numerous [13]. The proposed approach is a trade-off between the two ends of the spectrum and attempts to include relational dependencies between sensing modalities, while keeping it tractable for practical application. A hierarchical approach ensures that composite patterns are identified only when its constituting units at the lower level have been observed. In the current framework we have considered relations taken only two at a time, but we propose to explore relations between higher order cliques as future work.

D. Feature Selection and Pattern Classification

Once the feature vectors are extracted from the observed sensor time series, the next step is to classify these patterns into different categories based on the particular application. However, the feature vector $\mathbf{p}_{ij}$ obtained from $\tilde{\Pi}_{ij}$ has the dimension of $|\mathcal{Q}_i| \times |\Sigma_j|$ which is still high if $|\mathcal{Q}_i|$ and $|\Sigma_j|$ are large. More importantly, many features in $\mathbf{p}_{ij}$ may be zero since some transitions never occur. Therefore, it is necessary to perform feature selection in order to find the most representative and discriminative features and speed up the pattern classification process. Many standard methods can be found in feature selection literature [15]; a simple method that selects the features with large inter-class separation and small intra-class variance is adopted in this paper. Advanced methods such as forward/backward selection [15] and mRMR feature selection [16] may further improve the classification result by selecting the most representative and discriminative features; however, it is a topic of future research and not the focus of this paper.

Pattern classification for personnel detection is posed as a two-stage problem, i.e., the training stage and the testing stage. The sensor time series data sets are divided into three groups: i) partition data, ii) training data, and iii) testing data. The partition data set is used to generate partition planes that are used in the training and the testing stages. The training data set is used to generate the training patterns of different classes for the pattern classifier. Multiple sets of training data are obtained from independent experiments for each class in order to provide a good statistical spread of patterns. Subsequently, the class labels of the testing patterns are generated from testing data in the testing stage. The partition data sets may be part of the training data sets, whereas the training data sets and the testing data sets must be mutually exclusive.

The partition data is wavelet-transformed with appropriate scales to convert the one-dimensional numeric time series data into the wavelet image. The corresponding wavelet surface is analyzed using the *maximum entropy principle* [14] to generate the partition planes that remain invariant for both the training and the testing stage. The scales used in the wavelet transform of the partitioning data also remain invariant during the wavelet transform of the training and the testing data. In the training stage, the wavelet surfaces are generated by transformation of the training data sets corresponding to different classes. These surfaces are symbolized using the partition planes to generate the symbol images. Subsequently, PFSAs (either D-Markov or xD-Markov machines) are constructed based on the corresponding symbol images, and the training patterns are extracted from these PFSAs. Similar to the training stage, the PFSA and the associated pattern is generated for different data sets in the testing stage.

Finally a classifier is trained using features of different classes extracted from training data and can be used to classify the features from test data set. There are plenty of choices available for design of both parametric and non-parametric classifiers in literature [17]. Among the parametric type of classifiers, one of the most common techniques is to consider up to two orders of statistics in the feature space. In other words, the mean feature is calculated for every class along with the variance of the feature space distribution in the training set. Then, a test feature vector is classified by using the Mahalanobis distance or the Bhattacharya distance of the test vector from the mean feature vector of each class. However, these methods lack in efficiency if the feature space distribution cannot be described by second order statistics (i.e., non-Gaussian in nature). In the present context, Gaussian feature space distribution cannot be ensured due to the non-linear nature of the partitioning feature extraction technique. Therefore, a non-parametric classifier, such as the k-Nearest Neighbors (k-NN) classifier may a better candidate for this study [17]; however, in general, any other suitable classifier, such as the Support Vector Machines (SVM) or the Gaussian Mixture Models (GMM) may also be used.

IV. Results of Field Data Analysis

Field data were collected in the scenario illustrated in Fig. 1. Multiple data runs were made to collect data sets of all three classes, i.e., no target, human, and animal. The data were collected over three days at different sites. A brief summary is given in Table I showing the number of runs of each class.

Each data set, sampled at a sampling frequency of $Fs = 10$ kHz, has 1×10^5 data points that correspond to 10 seconds of the experimentation time. In order to test the capability of the proposed algorithm in target detection, another group of data were collected with no target present. The problem of target detection is then formulated as a binary pattern classification, where the no target present data are considered as one class, and the others with target present (i.e., human or animal) are considered to belong to the other class. The data sets, collected by the channel of seismic sensors that are orthogonal to the ground surface and the PIR sensors that are collocated with the seismic sensors, are used for target detection and classification. For computational efficiency, the original seismic and PIR data were downsampled by a factor of 10.

Table I
THE NUMBER OF FEATURE VECTORS FOR EACH TARGET CLASS IN THE DATA SETS USED FOR THREE-WAY CROSS VALIDATION (SET 3)

	Day 1	Day 2	Day 3	Total
No target	50	28	32	110
Human	30	22	14	66
Animal	20	6	18	44

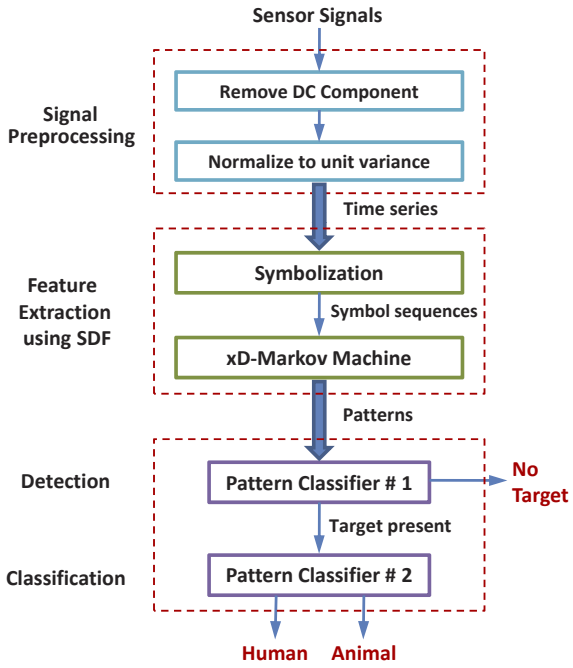


Figure 7. Flow chart of the problem of target detection and classification

Figure 7 depicts the flow chart of the proposed detection and classification algorithm that is constructed based on the theories of symbolic dynamic filtering (SDF) and k-nearest neighbors (k-NN) classifier [17]. The proposed algorithm consists of four steps, namely, signal preprocessing, feature extraction, detection, and classification, as shown in Fig. 7.

In the signal preprocessing step, the DC component of the seismic signal is eliminated, and the signal is normalized to unit variance. The amplitude of seismic signal of a horse with heavy payload passing by far away may be similar to that of a pedestrian passing by at close distance due to the fact that the SNR decreases rapidly with the distance between the sensor and the target. The normalization of all signals to unit variance makes the pattern classifier independent of the signal amplitude and any discrimination should be solely texture-dependent. For PIR signals, only the DC component is removed and the normalization is not performed because the range of PIR signals do not change.

In the feature extraction step, SDF captures the signatures of the preprocessed sensor time-series for representation as low-dimensional feature vectors. Based on the spectral analysis of the ensemble of seismic data at hand, a series of pseudo-frequency from the 1-20 Hz bands have been chosen to generate the scales for wavelet transform, because these bands contain a very large part of the footstep energy [6]. Similarly, a series of pseudo-frequency from the 0.2-2 Hz bands have been chosen for PIR signals to generate the scales. Upon generation of the scales, continuous wavelet transforms (CWT) are performed with appropriate wavelet basis function on the seismic and PIR signals. $db7$ is used for seismic signals since it matches the impulsive shape of seismic signals very well, $db1$ is used for the PIR case since PIR signals are close to square waves. A maximum-entropy wavelet surface partitioning is then performed. Selection of the alphabet size $|\Sigma|$ depends on the characteristics of the signal: a small alphabet size is robust against noise and environmental variation, while a large alphabet size has more discriminant power for identifying different objects. The same alphabet is used for both target detection and classification and the issues of alphabet size optimization and data set partitioning are not addressed in this paper. The execution of the code takes less than 1 second for SDF to process a data set of 1×10^4 points with the following choice of parameters: alphabet size $|\Sigma_1| = |\Sigma_2| = 8$, number of scales $|\alpha_1| = |\alpha_2| = 4$, for seismic and PIR sensor signals, respectively. $\mathcal{M}_{1\to 2}$ and $\mathcal{M}_{2\to 2}$ are used to form the composite pattern.

The next step is to perform pattern classification on the feature vectors. Two classifiers are needed as in the flow chart of Fig. 7: one for target detection to decide whether a target is present or not, and the other to identify the target. All classifiers are implemented by k-NN classifier. The available feature vectors are divided into two sets: a training set and a testing set. In the numerical results presented in the following sections, a three-way cross-validation [17] is used. The data is divided into three sets by date (Day 1, Day 2, Day 3) and three different sets of experiments are performed:

1) Training: Day 1 + Day 2; Testing: Day 3
2) Training: Day 1 + Day 3; Testing: Day 2
3) Training: Day 2 + Day 3; Testing: Day 1

Training and testing on feature vectors from different days is very meaningful in practice. In each run of cross-validation, no prior information is assumed for the testing site or testing

Table II
CONFUSION MATRICES OF THE THREE-WAY CROSS-VALIDATION RESULTS USING SEISMIC, PIR AND FUSION OF THE TWO SENSORS

Seismic Sensor	No target	Human	Animal
No target	76	5	29
Human	16	29	21
Animal	6	13	25

PIR Sensor	No target	Human	Animal
No target	110	0	0
Human	3	51	12
Animal	0	9	35

Sensor Fusion	No target	Human	Animal
No target	110	0	0
Human	1	52	13
Animal	0	1	43

Table III
COMPARISON OF THE DETECTION AND CLASSIFICATION ACCURACY BY USING SEISMIC, PIR AND FUSION OF THE TWO SENSORS

	Seismic	PIR	Fusion
Detection	74.5%	98.6%	99.5%
Classification	61.4%	80.4%	87.2%

data. The classifiers' capability to generalize to an independent data set is thoroughly tested in the three-way cross-validation.

Following Fig. 7, the following cases are tested:

1) Detection of target presence against target absence;
2) Classification of target type, i.e., Human vs. Animal.

Table II shows the confusion matrices of the three-way cross-validation using seismic sensor, PIR sensor and fusion of the two sensors. The shaded area in Table II represents the confusion matrices of target classification. Table III summaries the detection and classification accuracy in Table II. It is observed that the seismic sensor does not perform well for training and testing in different test sites. This is because seismic sensor is not site independent; variation in ground impedance and texture may affect the performance in target detection and classification. PIR sensors are almost site independent and achieve much higher accuracy than seismic sensors in both detection and classification. By using the composite pattern generated by fusing the signals from seismic and PIR sensors, the detection and classification results are further improved.

V. Summary, Conclusions and Future Work

This paper presents a feature-level fusion method for personnel detection using multimodal sensors. These features are extracted as statistical patterns by constructing xD-Markov machines from time series data collected from multimodal sensors. An appropriate selection of the basis function and the scale range allows the wavelet-transformed signal to be de-noised relative to the original noise-contaminated signal before partitioning of the resulting wavelet image for symbol generation. The xD-Markov machine identifies the cross-dependencies among different sensors and mitigates loss of significant information as compared to set-theoretic information fusion method. A distinct advantage of the proposed method is that the low-dimensional feature vectors, extracted from the xD-Markov machine, can be computed *in situ* and communicated in real time over a limited-bandwidth wireless sensor network with limited-memory nodes.

The proposed method has been validated on a set of field data collected from different locations on different days. A comparative evaluation is performed on the feature vectors extracted from single seismic and single PIR sensors as well as the composite pattern generated by fusion of the seismic and PIR sensors using xD-Markov machine. Results show that, while PIR sensors alone perform better than seismic sensors alone, the co-dependence-aware fusion further improves the detection and classification performance.

While there are many research issues that need to be resolved before exploring commercial applications of the proposed method, the following topics are under active research:

- Exploration of alternative ways for construction of relational PFSAs from wavelet images with mutiple scales;
- Improvement of the feature selection procedure by adopting more advanced methods [15], [16];
- Development of algorithms to extract relational dependencies among three or more symbol sequences;
- Comparative evaluation of the proposed sensor fusion method with Dempster-Shafer and Bayesian network approaches [17] at the decision fusion level.

Acknowledgement

This work has been supported in part by the U.S. Army Research Laboratory and the U.S. Army Research Office under Grant No. W911NF-07-1-0376. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the sponsoring agencies.

The authors would like to thank Dr. Abhishek Srivastav and Mr. Soumik Sarkar for their assistance in developing and implementing the xD-Markov machine.

References

[1] G. L. Goodman, "Detection and classification for unattended ground sensors." in *Proceedings of Information Decision and Control 99*, 1999, pp. 419–424.

[2] D. Li, K. Wong, Y. H. Hu, and A. Sayeed, "Detection, classification, and tracking of targets," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 17 –29, March 2002.

[3] Y. Tian and H. Qi, "Target detection and classification using seismic signal processing in unattended ground sensor systems," in *International Conference on Acoustics Speech and Signal Processing*, 2002.

[4] G. P. Succi, D. Clapp, R. Gampert, and G. Prado, "Footstep detection and tracking," in *Unattended Ground Sensor Technologies and Applications III*, vol. 4393, no. 1. SPIE, 2001, pp. 22–29.

[5] J. Lacombe *et al.*, "Seismic detection algorithm and sensor deployment recommendations for perimeter security," in *Unattended Ground, Sea, Air Sensor Technologies & Applications VIII*, vol. 6231. SPIE, 2006.

[6] K. M. Houston and D. P. McGaffigan, "Spectrum analysis techniques for personnel detection using seismic sensors," in *Unattended Ground Sensor Technologies & Applications V*, vol. 5090. SPIE, pp. 162–173.

[7] Z. Zhang, X. Gao, J. Biswas, and K. K. Wu, "Moving targets detection and localization in passive infrared sensor networks," in *Proceedings of 10th International Conference on Information Fusion*, July 2007.

[8] T. Damarla and D. Ufford, "Personnel detection using ground sensors," in *Unattended Ground, Sea, and Air Sensor Technologies and Applications IX*, vol. 6562, no. 1. Orlando, FL, USA: SPIE, 2007, p. 656205.

[9] H. Xing, F. Li, H. Xiao, Y. Wang, and Y. Liu, "Ground target detection, classification, and sensor fusion in distributed fiber seismic sensor network," in *Advanced Sensor Systems and Applications III*, vol. 6830. SPIE, 2008, p. 683015.

[10] A. Gunatilaka and B. Baertlein, "Feature-level and decision-level fusion of noncoincidently sampled sensors for land mine detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 577–589, 2001.

[11] A. Ray, "Symbolic dynamic analysis of complex systems for anomaly detection," *Signal Processing*, vol. 84, no. 7, pp. 1115–1130, 2004.

[12] S. Gupta and A. Ray, "Symbolic dynamic filtering for data-driven pattern recognition," in *Pattern Recognition: Theory and Application*. Nova Science Publishers, Hauppage, NY, 2007, ch. 2, pp. 17–71.

[13] S. Sarkar, D. S. Singh, A. Srivastav, and A. Ray, "Semantic sensor fusion for fault diagnosis in aircraft gas turbine engines," in *Proceedings of 2011 American Control Conference*, San Francisco, CA, 2011.

[14] X. Jin, K. Mukherjee, S. Gupta, and A. Ray, "Wavelet-based feature extraction using probabilistic finite state automata for pattern classification," *Pattern Recognition, in press*, 2010.

[15] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, March 2003.

[16] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226 –1238, August 2005.

[17] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.